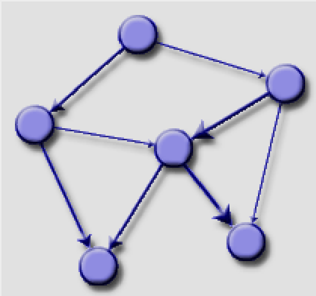
Causal Modeling and Machine Learning
Beijing, China, June 2014



Estimation of causal direction in the presence of latent confounders and linear non-Gaussian SEMs

Shohei Shimizu
Osaka University, Japan

with

Kenneth Bollen
University of North Carolina, Chapel Hill, USA

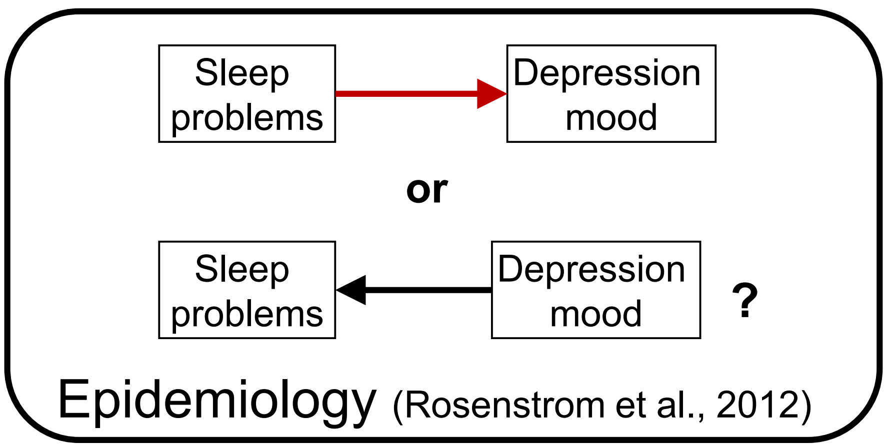
Abstract

- Estimation of causal direction of two observed variables in the presence of latent confounders
- A key challenge in causal discovery
- Propose a non-Gaussian method
- Not require to specify the number of latent confounders
- Experiments on artificial and sociology data

Background

Motivation

- Causality is a main interest in many empirical sciences
- Many recent methods for estimating causal directions (with no temporal information)
  - Linear non-Gaussian model (Dodge & Rousson 2001; Shimizu et al., 2006)
  - Nonlinear model (Hoyer et al., 2009; Zhang & Hyvarinen, 2009; Peters et al. 2011)



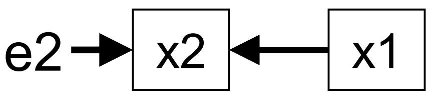
Which is dominant?

- Another important challenge: Latent confounders

Structural equation modeling (SEM) (Bollen, 1989; Pearl, 2000, 2009)

- A framework for describing causal relations
- An example (of linear cases):

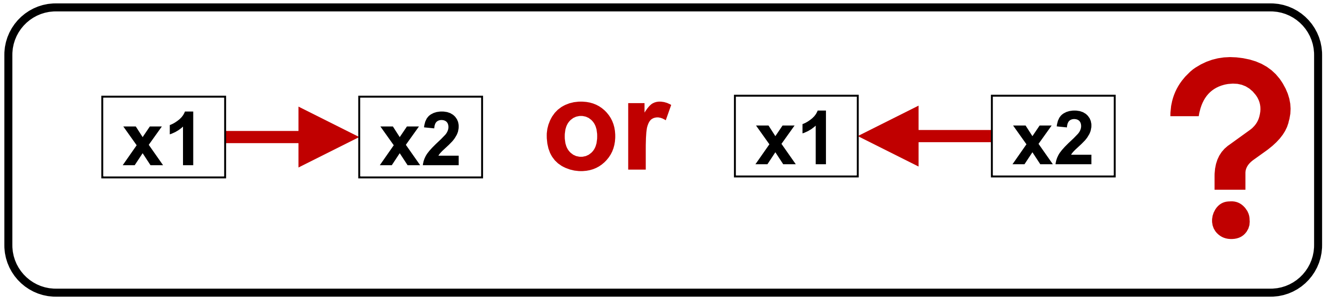
$$\begin{aligned}\boldsymbol{x_2} &:= \boldsymbol{f(x_1, e_2)} \\ &= \boldsymbol{b_{21} x_1 + e_2}\end{aligned}$$



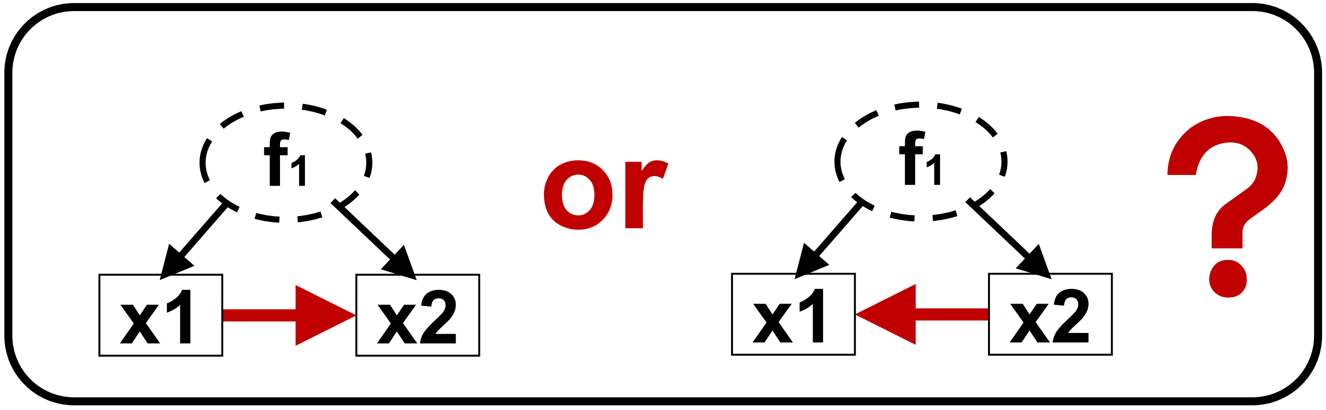
  - The value of x_2 is determined by the values of x_1 and error/exogenous variable e_2 through the linear function
- Generally speaking, if the value of x_1 is changed and that of x_2 also changes, then x_1 causes x_2

Major challenges

1. Estimation of causal direction when temporal information is not available



2. Coping with latent confounders



Non-Gaussian approach: LiNGAM

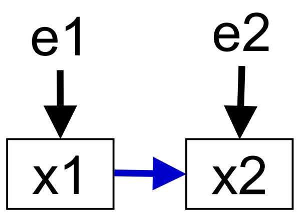
(Linear Non-Gaussian Acyclic Model) (Shimizu et al., 2006)

- Acyclic SEMs with different directions distinguishable
  (Dodge & Rousson, 2001; Shimizu et al., 2006)

Model 1:

$$x_1 = e_1$$

$$x_2 = b_{21} x_1 + e_2$$

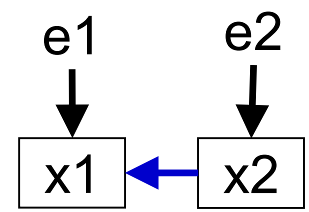


or

Model 2:

$$x_1 = b_{12} x_2 + e_1$$

$$x_2 = e_2$$



where e_1 and e_2 are error/exogenous variables

- Fundamental assumptions:
  - e1 and e2 are non-Gaussian
  - Independence btw. e1 and e2 (No latent confounders)

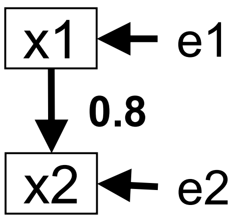
Different directions give different data distributions

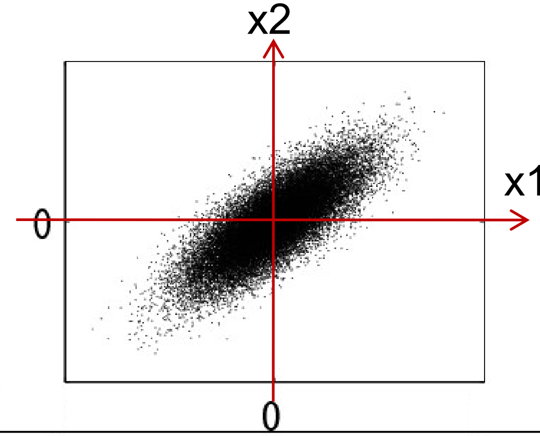
Gaussian | Non-Gaussian (uniform)

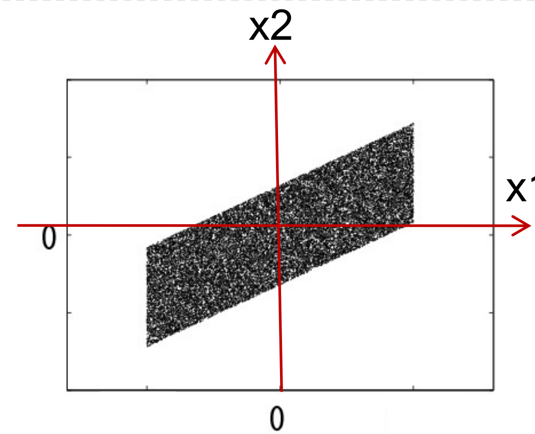
Model 1:

$$x_1 = e_1$$

$$x_2 = 0.8x_1 + e_2$$



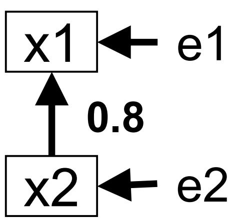


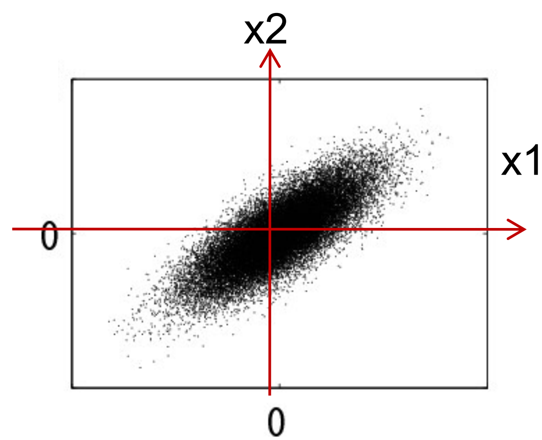


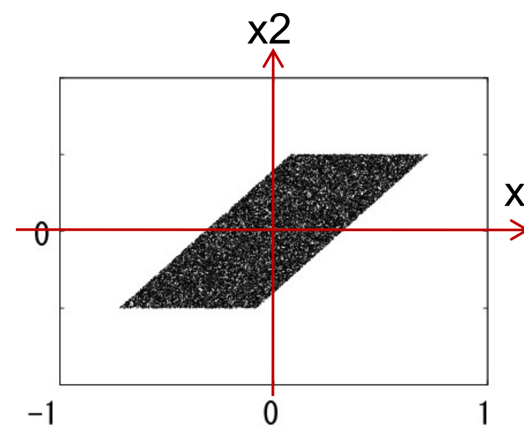
Model 2:

$$x_1 = 0.8x_2 + e_1$$

$$x_2 = e_2$$







$$E(e_1) = E(e_2) = 0,$$

$$\mathrm{var}(x_1) = \mathrm{var}(x_2) = 1$$

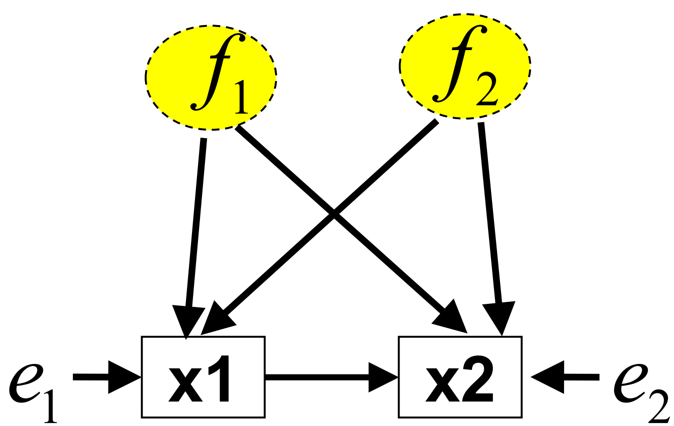
LiNGAM with latent confounders

(Hoyer, Shimizu & Kerminen, 2008)

- Extension to incorporate non-Gaussian latent confounders f_q

$$\boxed{x_i = \mu_i + \sum_{q=1}^{Q} \lambda_{iq} f_q + \sum_{j \neq i} b_{ij} x_j + e_i}$$

where, WLG, $f_q \ (q = 1, \cdots, Q)$ are independent:

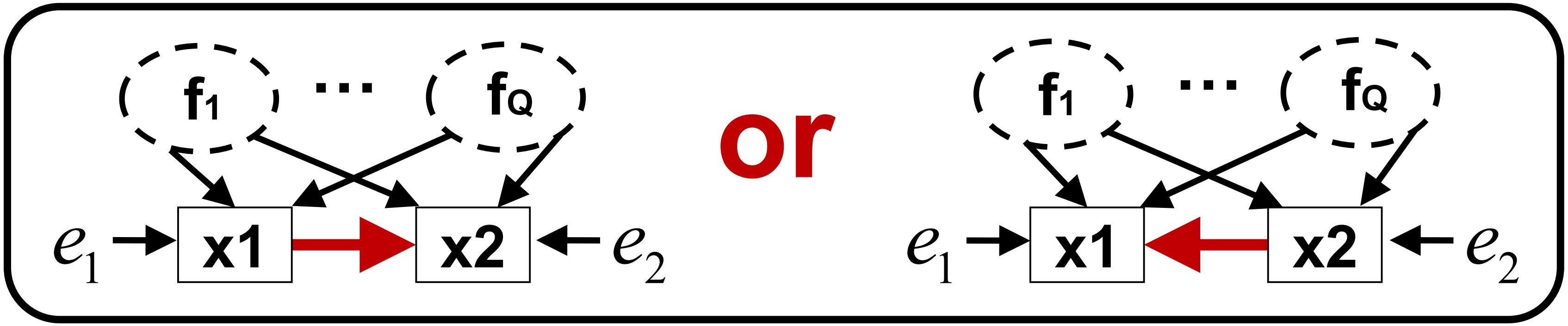


$$x_1 = \mu_1 + \sum_{q=1}^{Q} \lambda_{1q} f_q + e_1$$

$$x_2 = \mu_2 + \sum_{q=1}^{Q} \lambda_{2q} f_q + b_{21} x_1 + e_2$$

Previous estimation approaches

- Explicitly model latent confounders and compare two models with opposite directions of causation
  - Maximum likelihood principle (Hoyer et al., 2008)
  - Bayesian model selection (Henao & Winther, 2011)
  - Laplace / finite mixture of Gaussians for p(e_i)
- Require to specify the number of latent confounders, which is difficult in general



Our proposal

Reference:

Shimizu and Bollen (2014)
Journal of Machine Learning Research

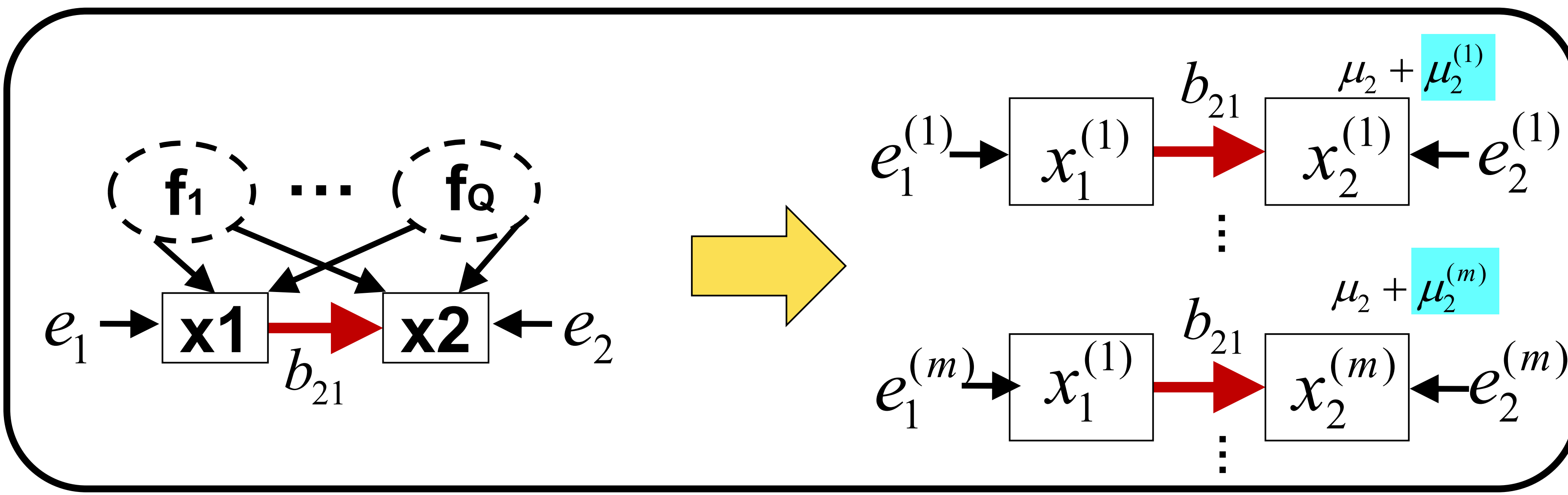
In press

Key idea (1/2)

- Another look at the LiNGAM with latent confounders:

***m*-th obs.:** $$x_2^{(m)} = \mu_2 + \underbrace{\sum_{q=1}^{Q} \lambda_{2q} f_q^{(m)}}_{=:\ \mu_2^{(m)}} + b_{21} x_1^{(m)} + e_2^{(m)}$$

Observations are generated from the LiNGAM model with possibly different intercepts $\mu_2 + \mu_2^{(m)}$



Key idea (2/2)

- Include the sums of latent confounders as the observation-specific intercepts:

***m*-th obs.:** $$x_2^{(m)} = \mu_2 + \underbrace{\boxed{\sum_{q=1}^{Q} \lambda_{2q} f_q^{(m)}}}_{=:\ \mu_2^{(m)}} + b_{21} x_1^{(m)} + e_2^{(m)}$$

Obs.-specific intercept

- Not explicitly model latent confounders
- Neither necessary to specify the number of latent confounders Q nor estimate the coefficients λ_{2q}

Our approach

- Compare these two LiNGAM models with opposite directions:

Model 3 (x1 → x2)

$$x_1^{(m)} = \mu_1 + \boxed{\mu_1^{(m)}} + e_1^{(m)}$$
$$x_2^{(m)} = \mu_2 + \boxed{\mu_2^{(m)}} + b_{21}x_1^{(m)} + e_i^{(m)}$$

Model 4 (x1 ← x2)

$$x_1^{(m)} = \mu_1 + \boxed{\mu_1^{(m)}} + b_{12}x_2^{(m)} + e_1^{(m)}$$
$$x_2^{(m)} = \mu_2 + \boxed{\mu_2^{(m)}} + e_2^{(m)}$$

- Many additional parameters $\boxed{\mu_i^{(m)}}$ $(i = 1,2; m = 1,\cdots,n)$
- Prior for the observation-specific intercepts $\boxed{\mu_i^{(m)}}$
- Other para. low-informative: Gaussian with large sd.
- Bayesian model selection (marginal likelihoods)

Prior for the observation-specific intercepts $\boxed{\mu_1^{(m)}} = \sum_{q=1}^{Q} \lambda_{1q} f_q^{(m)}, \quad \boxed{\mu_2^{(m)}} = \sum_{q=1}^{Q} \lambda_{2q} f_q^{(m)}$

- Motivation: Central limit theorem
  - Sums of independent variables tend to be more Gaussian
- Approximate the density by a bell-shaped curve dist.

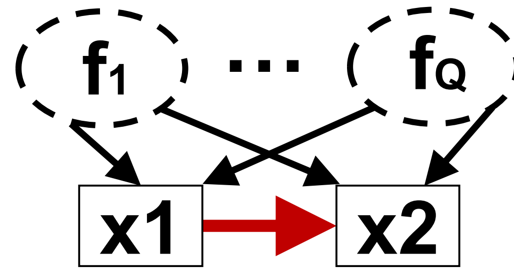
$$\begin{bmatrix} \mu_1^{(m)} \\ \mu_2^{(m)} \end{bmatrix} \sim$$ ***t*-distribution with sd** σ_1, σ_2 **,**
correlation σ_{12}**, and DOF** ν

- Select the hyper-parameter values that maximize the marginal likelihood: Empirical Bayes
  - $\sigma_l \in \{0, 0.2 \times \mathrm{sd}(x_l), \cdots, 1.0 \times \mathrm{sd}(x_l)\}, \quad \sigma_{12} \in \{0, \pm 0.1, \cdots, \pm 0.9\}$
  - DOF ν fixed to be 6 in the experiments below
- Small σ_l means similar intercepts

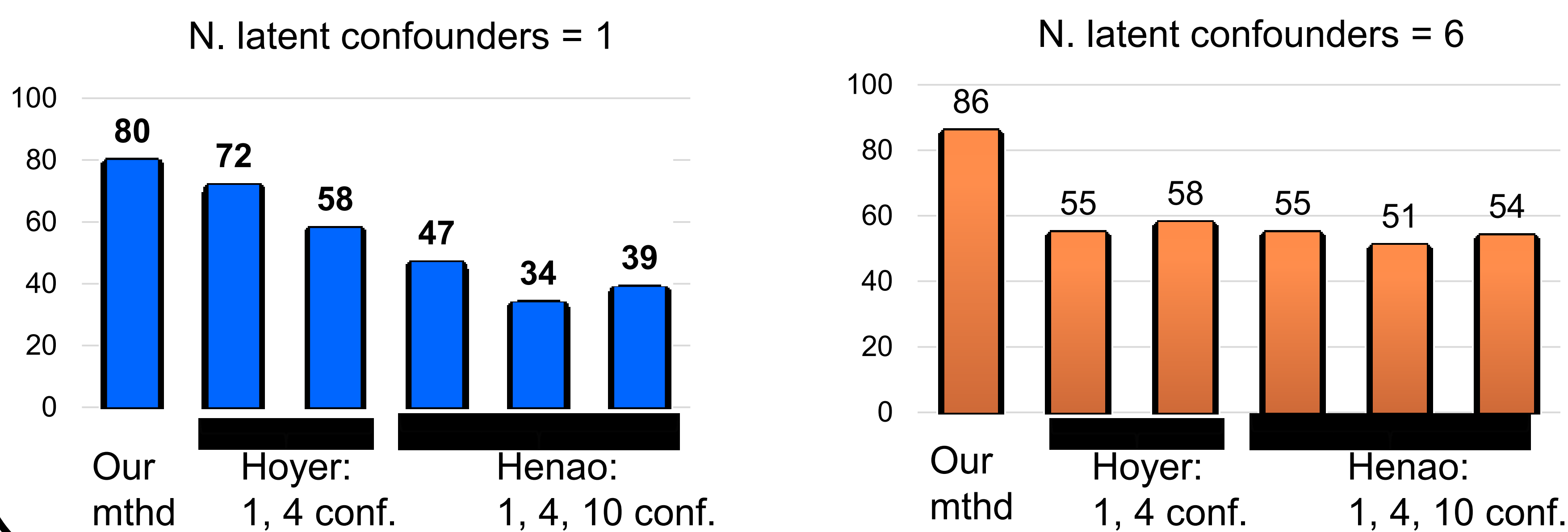
Experiments on artificial data

Experimental results (100 obs.)

- Data generated from LiNGAM with latent confounders
- Various non-Gaussian distributions
  - Laplace, Uniform, asymmetric dist. etc.
- Our method uses Laplace for p(e_i)



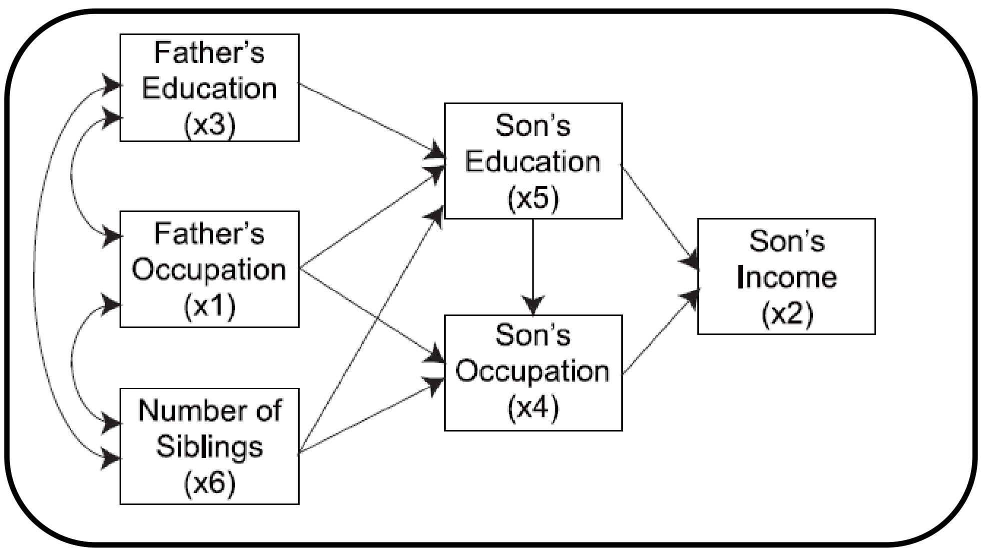
Numbers of successful discoveries (100 rep.)



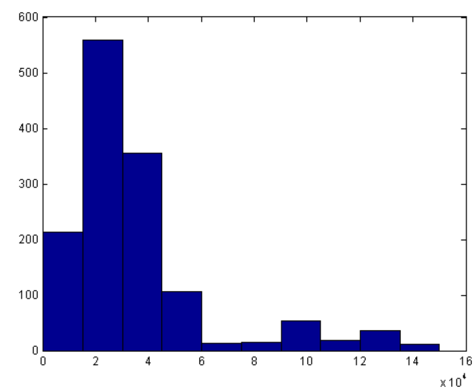
Experiment on sociology data

Sociology data

- Source: General Social Survey (n=1380)
  - Non-farm background, ages 35-44, white, male, in the labor force, no missing data for any of the covariates, 1972-2006

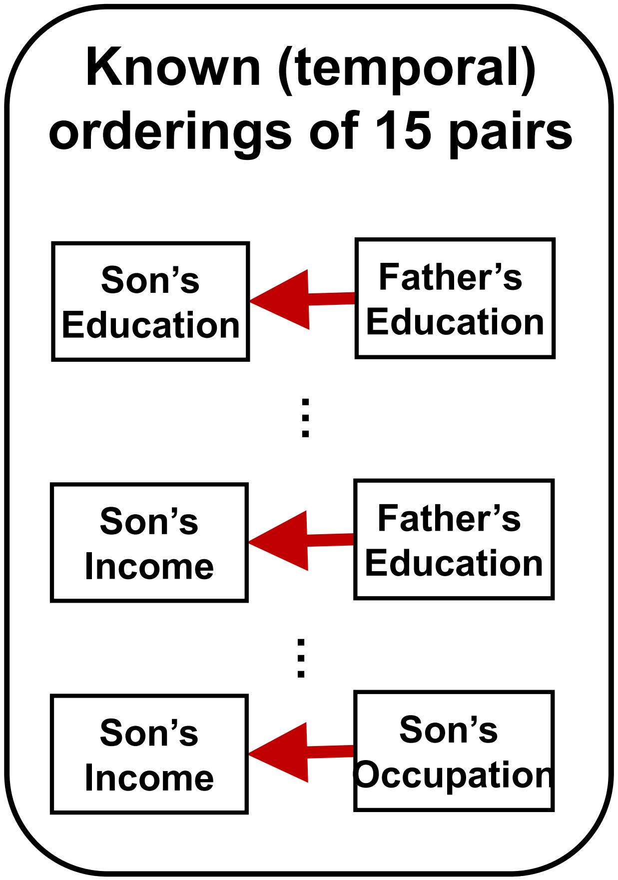


Status attainment model
(Duncan et al., 1972)



x2: Son's Income

Evaluation of our method using the sociology data



Possible Directions	Our method	Hoyer (2008) 1 conf.	Hoyer (2008) 4 conf.	Henao (2011) 1 conf.	Henao (2011) 4 conf.	Henao (2011) 10 conf.
FO <- FE	✔		✔			✔
SI <- FO	✔				✔	✔
SI <- FE	✔			✔	✔	✔
SI <- SO	✔			✔	✔	
SI <- SE	✔	✔	✔	✔		✔
SI <- NS	✔	✔	✔			
SO <- FO	✔	✔	✔	✔	✔	✔
SO <- SE	✔	✔		✔	✔	✔
SO <- SE	✔	✔	✔			
SO <- NS	✔	✔	✔	✔		
SE <- FO				✔		
SE <- FE	✔	✔	✔	✔	✔	
SE <- NS	✔	✔	✔	✔		
NS <- FO		✔				✔
NS <- FE		✔	✔		✔	✔
N. successes	12	10	9	9	7	8
Precisions	0.80	0.67	0.60	0.60	0.47	0.53

Conclusions

Conclusions

- Estimation of causal direction in the presence of latent confounders is a major challenge in causal discovery
- Our proposal: Fit linear non-Gaussian SEM with possibly different intercepts to data
- Future works
  - Test other informative priors for observation-specific intercepts
  - Implement a wider variety of error/prior distributions (e.g., learn DOF of t dist.)
  - Develop extensions using nonlinear/cyclic models (Hoyer et al., 2009; Zhang & Hyvarinen, 2009; Lacerda et al., 2008) instead of LiNGAM